DOCUMENT RESUME

ED 045 014 HE 001 834

AUTHOR	Sharon, Amiel T.
TITLE	Measurement of College Achievement by the College-Level Examination Program.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-70-38; RDR-69-70-NO-5
PUB DATE	Jun 70
NOTE	17p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.95
DESCRIPTORS	*Academic Achievement, Achievement Tests, *Higher Education, Research, *Tests, *Test Validity
IDENTIFIERS	*College-Level General Examinations

ABSTRACT

This report discusses the criterion-related and construct validity of the College-Level Examination Program's General Examinations primarily in terms of research studies conducted at institutions of higher education. While most of the research provides support for the validity of the examinations as measures of academic achievement in college, the results of many of the studies have alternative explanations. Correlations between the GE's and college grades obtained concurrently are moderately positive, but the validities of the tests for predicting success in upper-level studies are significantly lower. The GE's are also correlated positively with the amount of previous college experience. Significant gains on the tests are generally made by students over the first 2 years of college and the highest scores on each test are obtained by students intending to major in the subject covered by the test. The GE's relationship to other tests is also discussed, as well as their appropriateness for adults. (AF)

ED0 45014

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS

RDR-69-70, NO. 5

•

RESEARCH BULLETIN
RB-70-38 JUNE 1970

Measurement of College Achievement by the College-Level Examination Program

•

Amiel T. Sharon

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HE 001 834

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

ED 045014

MEASUREMENT OF COLLEGE ACHIEVEMENT BY THE COLLEGE-LEVEL EXAMINATION PROGRAM

Amiel T. Sharon

Educational Testing Service

Abstract

The criterion-related and construct validity of the College-Level Examination Program's General Examinations are discussed primarily in terms of research studies conducted at institutions of higher education. While most of the research provides support for the validity of the examinations as measures of academic achievement in college, the results of many of the studies have alternative explanations. The examinations correlate positively with course grades and amount of previous college instruction. Significant gains on the tests are generally made by students over the first two years of college and the highest scores on each test are obtained by students intending to major in the subject covered by the test. There is some question as to whether each of the five examinations is measuring a unique factor.

MEASUREMENT OF COLLEGE ACHIEVEMENT BY THE COLLEGE-LEVEL EXAMINATION PROGRAM

Amiel T. Sharon

Educational Testing Service

A large segment of the American population continues its education outside of school following termination of formal study. More than 82 million adult Americans are expected to be involved in educational programs outside the traditional school system by 1976 (American Institutes for Research, 1970). Most of these "students" are not pursuing academic degrees but have more immediate vocational objectives. Their learning activities are conducted by business, government, unions, military services, correspondence schools, antipoverty programs, community organizations, and instructional television.

The ever-increasing need for college graduates is encouraging many adults with nontraditional educational backgrounds to consider undertaking formal schooling which would lead them to a college degree. One way in which such people can demonstrate their previous educational achievements is by taking the General Examinations (GEs) of the College-Level Examination Program (CLEP).[1]

The GEs are intended to provide a comprehensive measure of undergraduate achievement in five basic areas of liberal arts: English, natural sciences, humanities, mathematics, and social sciences-history. The tests are not designed to measure advanced training in any specific discipline but rather to assess a student's knowledge and comprehension of basic facts, concepts, and principles in each of the five subjects. The content covered by the GEs is similar to the content included in the program of study required of

many liberal arts students in the first two years of college. It has been developed by committees of specialists in each of the subject-matter fields. The committees work with test specialists in defining the topics to be covered, reviewing the test specifications, and suggesting and reviewing test questions.

In addition to being used for granting college credit or placement for military service experiences, television and correspondence courses, and independent study, the GEs are used for a variety of other purposes at collegiate institutions. They are employed for guiding students into appropriate curricula of study; admitting and placing transfer students; assessing student growth in various curricula; and selecting students for upper division studies. Many colleges and universities are also using the examinations for self-study, to research specific questions about types of students, courses, or curricula. The questions which are asked range from "How do our sophomores compare with those at other colleges in terms of their liberal arts education?" to "Does exposure to our liberal arts courses result in greater knowledge as measured by these tests?"

The most common procedure for demonstrating the appropriateness or validity of achievement tests, such as the GEs, is by means of content validation. The test content is developed systematically to be representative of the subject matter to be measured. In addition, empirical procedures such as item analysis aid the test specialists in deciding on which items to include in the examinations. Since the GEs have been constructed by rigorous procedures of content validation described elsewhere (ETS, 1965), the present report focuses on the empirical validity of the tests.

Two different types of empirical validity will be discussed: criterion-related validity and construct validity. Criterion-related validity is useful for prediction of future performance and assessment of current achievement level. The criterion-related validity of the GEs will be described in terms of the relationship of the tests to college grades. Although the grade-point average (GPA) criterion has been criticized for being unstable (Humphreys, 1968) and for failing to reflect certain desirable types of student traits such as ethicality, open-mindedness, altruism, maturity, and self-insight (Davis, 1964), its ready availability has promoted its use as a criterion of college success by many researchers.

Unlike criterion-related validity, construct validity aims to increase understanding of the educational or psychological attributes measured by a test. It requires the gathering of information from a variety of sources. The construct validity of the GEs will be described by the effect of college instruction on test performance and by the differential performance of various types of students on the examinations. The possibility of the examinations being inappropriate to certain types of students, a topic closely related to validity, will also be discussed.

Criterion-Related Validity

Positive correlations between the GEs and overall GPA, in most cases overall sophomore GPA, have been reported in studies conducted at six universities (Beanblossom, 1969b; College Board Validity Study Service, 1967, 1969; Fujuta, 1965; Goolsby, 1966; Schnitzen, 1969). Since GPA and the scores on GEs were collected simultaneously in these studies, these correlations represent the concurrent validity of the examinations. Invariably the English Composition

Test was found to be the most valid one, with a median coefficient of .46. The rank order of the validity coefficients of the four other examinations was not consistent across the different studies. Median validities were Natural Sciences .40, Humanities .40, Social Sciences-History .36, and Mathematics .30. These correlations indicate that there is a moderately positive, but far from perfect, relationship between the tests' scores and grades. This result is not too surprising since grades in many courses are based on objective tests similar in content and format to the GEs. Nevertheless, these results suggest that the tests can be used legitimately for granting course credit or placement in college.

The correlations between the GEs and grades in subjects corresponding to each test are in general no higher than the tests' correlations with overall GPA. This conclusion is based on studies conducted at two universities (Beanblossom,1969b; Goolsby, 1966). A probable explanation of these results is that overall GPA is more reliable than subject GPA because it is based on a larger number of courses.

The validity of the GEs when taken at the end of the sophomore year, for predicting junior or junior/senior grades, is significantly lower than the concurrent validity of the tests. Median validity coefficients computed on the basis of three studies (College Board Validity Study Service, 1969; Goolsby, 1966; Harris, 1968) were English Composition .36, Humanities .28, Natural Sciences .27, Social Sciences-History .26, and Mathematics .15. Again, the English Composition and the Mathematics Tests appear to be the most and least valid tests respectively. The reason for the low validity of the Mathematics Test may be that mathematics plays a very minor role in courses taught in the

last two years of college. The finding that the predictive validities of the GEs are lower than their concurrent validities indicates that the tests are less useful for guidance or prediction of success in upper-level studies than they are as measures of current achievement level.

Construct Validity

Construct validity indicates the extent to which a test can be said to measure a trait or a theoretical construct. It also refers to the ability of a test to yield reasonable results, consistent with expectations. For example, a scholastic achievement test should yield higher scores for those who have more education than for those who have less education; history majors should score higher on a history test than biology majors; and students should have higher scores on an algebra test after taking an algebra course than before taking the course.

There are two reasonable expectations or implicit assumptions underlying the College-Level Examination Program which have implications for the construct validity of the GEs:

1. There is a gain in knowledge resulting from college instruction which can be measured by an examination.
2. The examinations employed to measure gain in knowledge are appropriate to the courses taught at the colleges.

These assumptions have implications which extend beyond those underlying the coefficient of correlation. In demonstrating that there is a positive correlation between test scores and grades no claim can be made that test scores or grades are affected by instruction. In order to determine whether a change in test performance is influenced by college instruction, it is

necessary to administer the test before and after the course of instruction. Also required would be the testing of one or more control groups (to which students would be randomly assigned) who would not receive instruction appropriate to the test or any instruction at all. Without a control group, any gains achieved on the examinations could be interpreted as resulting from intellectual growth rather than from a specific course of study. Unfortunately, it is difficult to have control groups in educational research. The notion of "manipulating" the learning of students for the sake of research is anathema to many educators. None of the studies which employed a "before-after" design to study score gains on the GEs employed a control group.

Harris and Booth (1969) reported on gains made on the GEs from the first to the sixth quarter by a group of 177 students who had taken the test twice. The mean gains ranged from a high of .6 of a standard deviation for the Social Sciences-History Test to a low of .3 of a standard deviation for the Mathematics Test. In relating the gains made on the GEs to grades in the courses corresponding to each test different results were found for the five tests. Students with higher grades achieved greater gains on the Humanities, Natural Sciences, and Social Sciences-History Tests only. The authors conclude that "on the average the better students in the various courses come into those courses with better scores on the respective tests and show greater gains" (p. 5). French (1965) described mean gains on the five examinations for a group of 81 students. These gains are similar in pattern and magnitude to those reported by Harris and Booth. Koby (1969) related gains to relevant course experiences for a sample of 82 students tested twice. Significant gains were made by the students only on the English Composition and Natural Sciences Tests.

The score gains reported in the three foregoing studies do not necessarily indicate that a particular college has done a good job or a poor job. The GEs are designed to cover subject matter content as taught at different colleges with different curricula, methods, and materials. They do not necessarily reflect all the objectives and emphases of any one college. In addition, the lack of control groups makes it difficult to know whether the score gains were a result of instruction or simply a result of maturation or intellectual growth occurring within the first two years of college.

The relationship of the GEs' scores to amount of previous instruction in a subject generally provides support for the validity of the examinations as measures of academic achievement. A relationship, however, does not prove cause, and thus it cannot conclusively demonstrate that the scores are affected by instruction. Nevertheless, a lack of relationship between the GEs' scores and amount of previous instruction would have led one to question the validity of the tests.

Beanblossom (1969b) correlated three GEs with the number of college credits taken in corresponding subjects. He concluded on the basis of his results that exposure to liberal arts courses "definitely" results in greater knowledge in natural sciences, "to some extent" in humanities, and "hardly at all" in social sciences and history. Selective factors, however, such as students taking more courses in their strong subjects, could account for these results.

The expectation that the tests' scores increase with the amount of formal college education completed has been confirmed by an analysis of the scores of 44,000 servicemen tested through the United States Armed Forces Institute (College Entrance Examination Board, 1968). There appears to be a steady and

significant progression of scores on all tests from those who have completed high school to those who have completed four years of college. Servicemen completing four years of college score about one standard deviation higher on each of the examinations than those who have not attended college. Similar results have been reported by Fagin (1969). She found a significant relationship between formal educational level and test performance for a group of 319 women. It should not be inferred from these two studies that the higher scores are necessarily the result of college study. It may be equally plausible to assume that individuals tend to remain longer in college because they perform well on tests.

The relationship of amount of high school preparation to the tests' scores was determined with the national freshman norming sample consisting of about 2500 second-term college students (Haven, 1967). Although the examinations were not intended to measure high school achievement, scores on all tests correlated positively with the number of years of appropriate course work completed in high school.

Additional results relating to the construct validity of the examinations have emerged from the data collected with the national norming sample of approximately 2600 college sophomores (Haven, 1964). The scores of sophomores intending to major in different fields fell into expected patterns. The highest mean score on each of the five examinations was obtained by students intending to major in the field corresponding to the examination. For example, those intending to major in social sciences performed best on the Social Sciences-History Test while those majoring in humanities or fine arts scored highest on the Humanities Test.

Relationships to Other Tests

The correlations found between the GEs and other standardized tests indicate that they have much in common with general aptitude and achievement measures. The correlations reported in almost all studies are between college entrance tests taken prior to admission to college and the GEs administered in the freshman or sophomore year. Because of changes taking place between the time of taking the entrance tests and the time of taking the GEs, the correlations reported are probably underestimates of the correlations that would have been obtained had the tests been taken by the students at the same time. While it is difficult to summarize the correlations because of the variety of tests used in the studies, correlations between the GEs and well-known standardized tests will be mentioned.

The English Composition Test was found to correlate .61 and .31 with the Scholastic Aptitude Test (SAT) Verbal and Mathematical sections respectively (Schnitzen, 1969). The corresponding correlations of the SAT with the Mathematics Test were .41 and .74. The correlation between the English Composition GE and the College Board English Composition Test was found to be .65 (Warren & Sylvan, 1969). A correlation of .70 was found between the combined score on the five GEs and the School and College Ability Tests (Goolsby, 1966).

The intercorrelation of the GEs indicate that to some extent all of the examinations except Mathematics are measuring the same ability or abilities (reading comprehension?). The median intercorrelations found in five studies ranged from a low of .12 between Humanities and Mathematics to a high of .56 between English Composition and Humanities. It should be pointed out, however, that the intercorrelations are much lower than expected of reliable tests

(above .9) measuring the same factors; thus, it is apparent that each test is also measuring some unique knowledge or skill.

Although the factorial composition of the GEs has not been determined, one could guess on the basis of the intercorrelations that two factors would account for most of the variance on the tests. The Mathematics Test would load high on a mathematical factor while the four other examinations would load high on a verbal factor.

Beanblossom (1969a) factor analyzed the scores from 11 precollege aptitude tests along with the scores of CLEP Social Sciences-History, Natural Sciences, and Humanities Tests. All three GEs loaded highly on a factor identified as a verbal factor. The Natural Sciences Test, unlike the other two examinations, also loaded highly on a factor identified as "general intelligence."

Appropriateness of the Tests for Adults

One of the major target populations of the College-Level Examination Program consists of mature adults who have not had any formal education in college. The content of the GEs, however, is based on the program of study offered to freshmen and sophomores attending liberal arts colleges who are mostly in their late teens. Does the content or the format of the examinations place the older candidates at a disadvantage?

An analysis of the scores of approximately 44,000 servicemen on the GEs appears to suggest that the tests are no more difficult for the older than for the younger examinees (College Entrance Examination Board, 1968). The oldest age group in this analysis, consisting of those of age 40 and over, was not the lowest scoring group on any of the examinations. In fact, this

group had the highest mean score of any age group on the Social Sciences-History and Humanities Tests. These two tests appear to be quite responsive to the accumulated value of life experience. The highest scores on the three other examinations occurred in the 22 to 24 age range. A limiting factor in the interpretation of this analysis is that the amount of formal education of servicemen at each age level was not known. While only 29 per cent of the sample had attended college, it is possible that the older age groups scored higher because they included more individuals with formal college education. Another possible explanation of the results is that the older servicemen in the sample were higher in ability or motivation as a result of self-selection.

French (1969) investigated the GEs' appropriateness with a sample of adult and black students. By using an inverse factor analysis on a matrix of the GEs' item responses he was able to identify 20 distinct hypothetical types of student, each defined by a certain set of items. Although the results suggest that the GEs do not give special advantage to any type of students, such as blacks or adults, it is difficult to have confidence in these results because the groups of subjects used was small and unrepresentative.

Unfortunately, there have been no studies on the comparative validity of the GEs for different types of students. If the relationship between the tests' scores and a criterion is different for various groups of examinees, then the tests may not be equally appropriate for all groups. It may be, for example, that speed is a relatively more important factor for adults than for younger persons, and it might consequently invalidate the tests as measures of achievement for adults.

Conclusion

In general, the research summarized provides support for the validity of the GEs as measures of academic achievement. Many of the studies reviewed, however, do not lead to definitive conclusions. Results showing score gains after course exposure and positive relationships between the tests and amount of previous instruction have alternative interpretations. Correlations between the GEs and college grades obtained concurrently are moderately positive, but the validities of the tests for predicting success in upper-level studies are significantly lower than their validities for assessing current achievement level. The research methodology for validating the GEs can be improved by employing criteria other than grades, by using control groups in score-gain studies, and by partialing out contaminating factors in correlational studies. Nevertheless, the relationships found between the GEs and certain relevant variables provide tentative support for the validity of the tests as measures of college-level achievement.

References

American Institutes for Research. Behavioral sciences newsletter for research planning, 1970, 7(5), 2.

Beanblossom, G. F. What do the CLEP General Examinations measure? Seattle: Bureau of Testing, University of Washington. Unpublished manuscript, 1969. (a)

Beanblossom, G. F. The use of CLEP scores in evaluating liberal arts curriculum. Seattle: Bureau of Testing, University of Washington. Unpublished manuscript, 1969. (b)

College Board Validity Study Service. Montclair State College CLEP Study. New York: College Entrance Examination Board. Unpublished manuscript, 1967.

College Board Validity Study Service. University of Delaware CLEP Study. New York: College Entrance Examination Board. Unpublished manuscript, 1969.

College Entrance Examination Board. Candidates tested through the United States Armed Forces Institute. New York: College Board, 1968.

Davis, J. A. Faculty perceptions of students, II. Faculty definition of desirable student traits. Research Bulletin 64-11. Princeton, N. J.: Educational Testing Service, 1964.

Educational Testing Service. ETS builds a test. Princeton, N. J., 1965.

Fagin, M. Analysis of the performance of selected women in Missouri on three examinations of the College Level Examination Program. St. Louis: University of Missouri, St. Louis. Unpublished manuscript, 1969.

French, J. W. Score gains on the Comprehensive College Tests during the first year. Sarasota, Florida: New College, Office of the College Examiner. Bulletin No. 10, 1965.

French, J. W. Types of students defined by items in the CLEP General Series of Achievement Tests. Sarasota: New College. Unpublished manuscript, 1969.

Fujuta, G. Preliminary report of a pilot study of the General Examinations of the Comprehensive College Tests, Spring 1965. Honolulu: Counseling and Testing Center, University of Hawaii. Unpublished manuscript, 1965.

Goolsby, T. M. The validity of a comprehensive college sophomore test battery for use in selection placement and advisement. Educational and Psychological Measurement, 1966, 26, 977-983.

Harris, J. An analysis of the College-Level Examination scores for the University of North Carolina at Greensboro. CEEB Southern Regional Meeting, 1968.

Harris, J., & Booth, E. An analysis of the performing of Georgia College freshmen and sophomores on the College-Level Examinations. Athens: Institute of Higher Education, University of Georgia. Unpublished manuscript, 1969.

Haven, E. W. The sophomore norming sample for the General Examinations of the College-Level Examination Program. Statistical Report 64-63. Princeton, N. J.: Educational Testing Service, 1964.

Haven, E. W. The freshman norms sample for the General Examinations of the College-Level Examination Program. Statistical Report 67-32. Princeton, N. J.: Educational Testing Service, 1967.

Humphreys, L. G. The fleeting nature of the prediction of college academic success. Journal of Educational Psychology, 1968, 59, 375-380.

Koby, H. L. Report on the general education program at Rio Grande College as suggested by the College-Level Examination Program. Rio Grande, Ohio. Unpublished manuscript, 1969.

Schnitzen, J. P. Descriptive statistics: College-Level Examination Program trial administration. Houston: Counseling and Testing Service, University of Houston. Unpublished manuscript, 1969.

Warren, J. R., & Sylvan, M. Relationship between the College-Level General Examination in English Composition and the English Composition Test. Santa Cruz: University of California. Unpublished manuscript, 1969.

Footnote

[1]The CLEP, which is sponsored by the College Entrance Examination Board, includes both the General and Subject Examinations.